

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Price, David H.

(ii) TITLE OF INVENTION: P-TEFb COMPOSITIONS, METHODS AND
SCREENING ASSAYS

(iii) NUMBER OF SEQUENCES: 68

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Arnold, White & Durkee
(B) STREET: P.O. Box 4433
(C) CITY: Houston
(D) STATE: TX
(E) COUNTRY: USA
(F) ZIP: 77210-4433

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US Unknown
(B) FILING DATE: Concurrently Herewith
(C) CLASSIFICATION: Unknown

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Fussey, Shelley P.M.
(B) REGISTRATION NUMBER: 39,458
(C) REFERENCE/DOCKET NUMBER: IOWA:012

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (512) 418-3000
(B) TELEFAX: (512) 418-3131

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1457 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 115..1326

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGTTGAGTCA ACAGCTGTAG ATACACCAAT TGTTGCCGAT TTCTTTCTTT TCGACTGTCG	60
GCTTCTCGCG AAACTGTGAT TGTGAAAATT GTACAAATAG AGGCAAATTT AACC ATG	117
Met	
1	
GCG CAC ATG TCC CAC ATG CTC CAG CAG CCT TCG GGG TCG ACG CCC TCC	165
Ala His Met Ser His Met Leu Gln Gln Pro Ser Gly Ser Thr Pro Ser	
5 10 15	
AAC GTG GGC TCC AGC TCA TCG CGC ACG ATG TCC CTG ATG GAG AAA CAA	213
Asn Val Gly Ser Ser Ser Arg Thr Met Ser Leu Met Glu Lys Gln	
20 25 30	
AAG TAC ATC GAG GAC TAC GAC TTT CCC TAC TGC GAC GAG AGC AAC AAA	261
Lys Tyr Ile Glu Asp Tyr Asp Phe Pro Tyr Cys Asp Glu Ser Asn Lys	
35 40 45	
TAC GAA AAG GTG GCG AAA ATT GGC CAA GGC ACC TTC GGA GAG GTT TTT	309
Tyr Glu Lys Val Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu Val Phe	
50 55 60 65	
AAG GCT CGC GAG AAA AAG GGC AAC AAG AAG TTT GTG GCC ATG AAG AAG	357
Lys Ala Arg Glu Lys Lys Gly Asn Lys Lys Phe Val Ala Met Lys Lys	
70 75 80	
GTG CTG ATG GAC AAC GAA AAG GAG GGC TTT CCC ATC ACG GCT CTG CGA	405
Val Leu Met Asp Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu Arg	
85 90 95	
GAG ATC CGC ATC CTG CAG CTG CTA AAG CAC GAG AAC GTG GTG AAT CTG	453
Glu Ile Arg Ile Leu Gln Leu Leu Lys His Glu Asn Val Val Asn Leu	
100 105 110	
ATC GAG ATC TGC CGC ACC AAG GCC ACC GCC ACG AAT GGT TAC AGA TCC	501
Ile Glu Ile Cys Arg Thr Lys Ala Thr Ala Thr Asn Gly Tyr Arg Ser	
115 120 125	
ACC TTC TAT TTG GTC TTT GAT TTC TGC GAA CAC GAT TTG GCA GGT CTT	549
Thr Phe Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly Leu	
130 135 140 145	
CTG TCC AAC ATG AAC GTC AAG TTC AGT CTG GGC GAG ATT AAG AAG GTT	597
Leu Ser Asn Met Asn Val Lys Phe Ser Leu Gly Glu Ile Lys Lys Val	
150 155 160	
ATG CAG CAG CTT TTA AAC GGT TTG TAT TAC ATC CAC AGC AAC AAG ATC	645
Met Gln Gln Leu Leu Asn Gly Leu Tyr Tyr Ile His Ser Asn Lys Ile	
165 170 175	

02251100-00000000

CTG CAC CGA GAC ATG AAA GCT GCC AAC GTG CTG ATT ACC AAG CAT GGC	693
Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Lys His Gly	
180 185 190	
ATC TTA AAG CTG GCT GAC TTT GGC TTG GCC CGT GCT TTT AGC ATT CCA	741
Ile Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Ile Pro	
195 200 205	
AAG AAC GAG AGT AAG AAT CGC TAT ACC AAT CGC GTA GTA ACC TTG TGG	789
Lys Asn Glu Ser Lys Asn Arg Tyr Thr Asn Arg Val Val Thr Leu Trp	
210 215 220 225	
TAC CGG CCG CCT GAG CTG CTA CTT GGT GAC CGC AAC TAT GGT CCA CCC	837
Tyr Arg Pro Pro Glu Leu Leu Leu Gly Asp Arg Asn Tyr Gly Pro Pro	
230 235 240	
GTG GAC ATG TGG GGA GCC GGC TGC ATA ATG GCC GAG ATG TGG ACA CGC	885
Val Asp Met Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr Arg	
245 250 255	
TCG CCC ATC ATG CAA GGC AAT ACG GAG CAG CAG CAG TTA ACC TTT ATT	933
Ser Pro Ile Met Gln Gly Asn Thr Glu Gln Gln Gln Leu Thr Phe Ile	
260 265 270	
TCG CAG CTA TGC GGC TCC TTT ACG CCG GAC GTG TGG CCG GGA GTG GAG	981
Ser Gln Leu Cys Gly Ser Phe Thr Pro Asp Val Trp Pro Gly Val Glu	
275 280 285	
GAG CTG GAG CTG TAC AAA TCC ATC GAG CTG CCA AAG AAC CAG AAG CGT	1029
Glu Leu Glu Leu Tyr Lys Ser Ile Glu Leu Pro Lys Asn Gln Lys Arg	
290 295 300 305	
CGA GTC AAG GAG CGC CTG CGT CCG TAT GTC AAG GAT CAA ACC GGC TGT	1077
Arg Val Lys Glu Arg Leu Arg Pro Tyr Val Lys Asp Gln Thr Gly Cys	
310 315 320	
GAT CTA TTG GAC AAA TTG CTG ACC CTT GAT CCC AAG AAA CGC ATC GAT	1125
Asp Leu Leu Asp Lys Leu Leu Thr Leu Asp Pro Lys Lys Arg Ile Asp	
325 330 335	
GCG GAC ACA GCT CTG AAT CAC GAC TTC TTC TGG ACG GAT CCC ATG CCC	1173
Ala Asp Thr Ala Leu Asn His Asp Phe Phe Trp Thr Asp Pro Met Pro	
340 345 350	
AGC GAC TTG AGC AAG ATG CTG TCC CAG CAC CTG CAG AGC ATG TTC GAG	1221
Ser Asp Leu Ser Lys Met Leu Ser Gln His Leu Gln Ser Met Phe Glu	
355 360 365	
TAC CTG GCG CAG CCA CGC CGC AGC AAC CAG ATG CGC AAC TAT CAC CAG	1269
Tyr Leu Ala Gln Pro Arg Arg Ser Asn Gln Met Arg Asn Tyr His Gln	
370 375 380 385	

CAA CTG ACC ACC ATG AAC CAG AAG CCC CAG GAC AAC AGT ATG ATT GAC 1317
Gln Leu Thr Thr Met Asn Gln Lys Pro Gln Asp Asn Ser Met Ile Asp
390 395 400

CGG GTT TGG TAGACTGCCA GAGGTGTACG CACCCGACTA ATAGTTTCTC 1366
Arg Val Trp

ACCTTCAACT AGCGTTAGGT TATTAGGTTA GTGTACAATA AAAATATTGG CATTGCATT 1426
AGCGCTTGCT CCAAATATAA AAAAAAAA A 1457

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala His Met Ser His Met Leu Gln Gln Pro Ser Gly Ser Thr Pro
1 5 10 15

Ser Asn Val Gly Ser Ser Ser Arg Thr Met Ser Leu Met Glu Lys
20 25 30

Gln Lys Tyr Ile Glu Asp Tyr Asp Phe Pro Tyr Cys Asp Glu Ser Asn
35 40 45

Lys Tyr Glu Lys Val Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu Val
50 55 60

Phe Lys Ala Arg Glu Lys Lys Gly Asn Lys Lys Phe Val Ala Met Lys
65 70 75 80

Lys Val Leu Met Asp Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu
85 90 95

Arg Glu Ile Arg Ile Leu Gln Leu Leu Lys His Glu Asn Val Val Asn
100 105 110

Leu Ile Glu Ile Cys Arg Thr Lys Ala Thr Ala Thr Asn Gly Tyr Arg
115 120 125

Ser Thr Phe Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly
130 135 140

Leu Leu Ser Asn Met Asn Val Lys Phe Ser Leu Gly Glu Ile Lys Lys
145 150 155 160

0395112888103692
Val Met Gln Gln Leu Leu Asn Gly Leu Tyr Tyr Ile His Ser Asn Lys
165 170 175
Ile Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Lys His
180 185 190
Gly Ile Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Ile
195 200 205
Pro Lys Asn Glu Ser Lys Asn Arg Tyr Thr Asn Arg Val Val Thr Leu
210 215 220
Trp Tyr Arg Pro Pro Glu Leu Leu Gly Asp Arg Asn Tyr Gly Pro
225 230 235 240
Pro Val Asp Met Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr
245 250 255
Arg Ser Pro Ile Met Gln Gly Asn Thr Glu Gln Gln Gln Leu Thr Phe
260 265 270
Ile Ser Gln Leu Cys Gly Ser Phe Thr Pro Asp Val Trp Pro Gly Val
275 280 285
Glu Glu Leu Glu Leu Tyr Lys Ser Ile Glu Leu Pro Lys Asn Gln Lys
290 295 300
Arg Arg Val Lys Glu Arg Leu Arg Pro Tyr Val Lys Asp Gln Thr Gly
305 310 315 320
Cys Asp Leu Leu Asp Lys Leu Leu Thr Leu Asp Pro Lys Lys Arg Ile
325 330 335
Asp Ala Asp Thr Ala Leu Asn His Asp Phe Phe Trp Thr Asp Pro Met
340 345 350
Pro Ser Asp Leu Ser Lys Met Leu Ser Gln His Leu Gln Ser Met Phe
355 360 365
Glu Tyr Leu Ala Gln Pro Arg Arg Ser Asn Gln Met Arg Asn Tyr His
370 375 380
Gln Gln Leu Thr Thr Met Asn Gln Lys Pro Gln Asp Asn Ser Met Ile
385 390 395 400
Asp Arg Val Trp

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4328 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CAGCCCTGCC	GACGGCCATA	CTTGAAAATA	CATTTTTTC	TGCAAAGTTT	GTCATTGTCA	60
CTGTGTGAAT	GGAATCTGTG	ATGTGTTGTG	GAATTAAAAAA	CGTCAAGTAA	ACAACCCGTA	120
ATGGTTAAAG	TGCACGGCGA	AAGCAGTGC	AATAACTATG	AATTGATACA	AAAGTTGCAT	180
AACACGTCGC	CTGGTGTGCG	GGTTAGTGTG	TTTTTGTCT	CGTTTCGTTT	CCGCCGCAGT	240
CGCAGTTTCC	AAAAAACCTC	ACCACACCAT	ACCATCTCCA	CCACGCACAC	ACACACACAA	300
ACAAACACGC	AGAGACGCGG	CGGCGGAAAA	AGTGTGCGGA	CCGCGGATTT	AACCCCTCGT	360
TCCAAACCCA	AATTGGAGTC	TCCCAAAAAC	AGCGAAATAT	CGAGTGTGGC	TTAGCCGATG	420
TGCCGTGCGA	TCCCCACTGC	CCCTTCCGTA	CCGCTGCCAC	CCCCGCCACA	GCAGCAACGC	480
ACACGGATAAC	GGACACAGAC	ACCAATACCA	GCGCACTCAA	GCACGGCCGA	CAAAGAAAGA	540
GCGCTCTCCC	TTCCCTTTG	TACAGTTAGT	TCCTACAGCT	GAATCAGCCA	AAAGAAATTA	600
CTAGGTCCAT	TCCGAGGCGC	AGTTTGCATG	TGAAACGGAG	GTCCCCGCAT	AACCACGCGG	660
AACCCGAAAT	TCCAGATCCC	CATCTCCGCT	GCACGGATAA	AGGAAACATA	CAACCATGAG	720
TCTCCTAGCC	ACGCCAATGC	CCCAGGCGGC	CACCGCCTCA	TCTTCTTCAT	CCGCCTCCGC	780
GGCCGCCTCG	GCCAGCGGGA	TTCCAATCAC	CGCCAACAAAC	AACCTGCCTT	TCGAGAAGGA	840
CAAGATCTGG	TACTTCAGCA	ACGATCAGCT	GGCCAATTG	CCAAGCAGAA	GATGCGGCAT	900
CAAGGGCGAC	GATGAGCTGC	AGTACCGCCA	GATGACCGCC	TATCTGATAC	AGGAAATGGG	960
TCAGCGTCTG	CAGGTGTCCC	AACTGTGCAT	CAACACGGCC	ATTGTGTACA	TGCATCGGTT	1020
CTACGCCTTT	CACTCCTTCA	CCCACTTCA	TCGCAACTCC	ATGGCGTCGG	CGAGCCTCTT	1080
CTTGGCCGCC	AAGGTAGAAG	AGCAACCGCG	GAAGCTGGAG	CATGTTATTC	GGGCCGCCAA	1140
CAAGTGCCTG	CCGCCGACCA	CCGAGCAGAA	TTACGCCAA	CTCGCCCAGG	AGCTTGTGTT	1200
CAACGAGAAC	GTGCTCCTGC	AGACGCTGGG	CTTCGATGTG	GCCATCGATC	ATCCGCACAC	1260
GCATGTGGTG	CGCACCTGCC	AGCTGGTCAA	AGCATGCAAG	GATCTGGCGC	AGACATCGTA	1320
CTTCTTGGCC	TCGAACAGCC	TGCATCTGAC	CTCGATGTG	CTCCAATATC	GCCCCACGGT	1380
CGTAGCCTGT	TTCTGCATTT	ACCTAGCCTG	CAAGTGGTCC	CGATGGGAGA	TCCCCCAGTC	1440
GACCGAGGGC	AAGCACTGGT	TCTACTATGT	GGACAAGACG	GTCTCGCTGG	ATTGCTAAA	1500

GCAGCTGACA GATGAGTTCA TCGCTATCTA TGAGAAGAGC CCGGCCGTC TGAAGTCTAA	1560
GCTTAACTCG ATCAAGGCGA TCGCCCAGGG AGCCAGCAAT CGGACAGCTA ACAGCAAGGA	1620
CAAACCAAAG GAGGACTGGA AGATCACCGA GATGATGAAG GGCTACCACT CAAACATCAC	1680
GACACCACCA GAGCTGTTAA ACGGCAACGA CAGCCGGAT CGGGACCGAG ATCGTGAACG	1740
GGAGAGAGAG CGGGAACGGG ATCCGTCGTC ACTACTGCCG CCACCGGCTA TGGTGCCGCA	1800
GCAAAGACGA CAGGATGGTG GACATCAGCG CTCGTCCTCA GTGAGCGGAG TGCCAGGCAG	1860
CAGCTCTCG TCGTCTTCCT CCAGTCACAA GATGCCAAAT TACCCCTGGTG GCATGCCGCC	1920
CGAAGCTCAT CGGGATCACA AGTCAAAGCA GCCGGGCTAT AACAAATCGAA TGCCCTCAAG	1980
TCACCAGCGT AGTAGTAGCA GTGGACTCGG TTCCTCGGA AGTGGCAGCC AGCACAGCAG	2040
CTCATCCTCG TCGTCTCAA GCCAGCAGCC TGGCCGACCG TCTATGCCCG TGGACTATCA	2100
CAAATCCTCT CGCGGCATGC CGCCGGTAGG CGTGGGCATG CCACCTCACG GCAGCCACAA	2160
GATGACTTCG GGCTCCAAGC CTCAACAGCC GCAGCAGCAG CCGGTCACAC ATCCATCCGC	2220
CTCTAATTCC TCTGCATCGG GCATGTCTC CAAGGATAAA TCCCAGAGCA ACAAAATGTA	2280
TCCGAACGCA CCGCCGCCAT ACAGTAATAG TGCCCTCAA AACCCGCTGA TGTCGCGTGG	2340
TGGATATCCA GGCCTAGCA ATGGATCCA GCCCCCGCCT CCCGCCGGAT ACGGCGGCCA	2400
TCGCAGCAA TCCGGCTCCA CCGTCCATGG CATGCCGCAT TTGAGCAGC AATTGCCCTA	2460
TTCCCAGAGC CAGAGCTACG GCCACATGCA GCAGCAGCCA GTGCCTCAGT CTCAGCAGCA	2520
ACAGATGCCT CCGGAGGCAT CCCAGCACTC GTTGCAGTCC AAGAACTCGC TCTTCAGTCC	2580
AGAGTGGCCA GACATTAAAA AGGAGCCCAT GTCGCAGTCG CAACCACAGC TTTTTAACGG	2640
TTTGCTACCC CCTCCTGCGC CTCCCGGCCA CGATTACAAG CTAAATAGCC ATCCGCGCGA	2700
CAAAGAAAGT CCCAAGAAAG AGCGACTAAC GCCAACCAAA AAGGATAAGC ACCGTCTGT	2760
AATGCCCTCA ATGGGCAGTG GGAACAGTTC CTCCGGCTCG GGATCATCAA AGCCGATGCT	2820
ACCGCCTCAC AAGAACGAGA TACCCCATGG CGGGGACCTG TTGACCAATC CTGGAGAGAG	2880
TGGAAGCCTA AAACGGCCA ACGAGATCTC GGGAAAGTCAG TATGGACTAA ATAAGCTGGA	2940
TGAAATAGAT AACAGTAATA TGCCTCGAGA AAAGCTTCGC AAGCTGGACA CTACAACCTGG	3000
ACTACCAACT TATCCGAATT ATGAGGAGAA ACACACGCCT CTGAATATGT CCAACGGAAT	3060
CGAGACAAACG CCGGATCTGG TGCGCAGTTT GCTAAAGGAG AGTCTGTGTC CATCGAACGC	3120

DRAFT - 2020-09-26

TTCGCTCCTG	AAACCGGATG	CCTTGACTAT	GCCTGGCCTG	AAACCACCGG	CCGAACTACT	3180
TGAGCCCATG	CCCGCACCAG	CGACAATCAA	GAAAGAACAG	GGAATAACTC	CGATGACCAG	3240
TTGGCTAGT	GGGCCGCAC	CCATGGATT	GGAAGTACCC	ACTAACAGG	CCGGAGAGAT	3300
TAAGGAGGAA	AGCAGCAGCA	AGTCCGAAAA	GAAAAAGAAG	AAGGATAAAC	ACAAACACAA	3360
GGAGAAGGAC	AAGTCCAAGG	ACAAGACGGA	AAAGGAGGAG	CGTAAGAACG	ACAAGAGGGA	3420
CAAGCAGAAG	GATCGTAGCG	GCAGCGGTGG	CAGCAAGGAC	AGTTCTCTTC	CCAATGAGCC	3480
TCTGAAGATG	GTTATCAAGA	ATCCCAACGG	CAGCCTGCAG	GCCGGTGCCT	CAGCTCCCCT	3540
TAAACTTAAG	ATCAGCAAAA	ATAAGGTTGA	ACCCAATAAC	TAATCTGCAG	CGGCGGGTCT	3600
GCCTGGCGCA	ATCGGATATG	GCTTGCCTCC	AACTACGGCT	ACCACCACAT	CCGCTTCGAT	3660
CGGAGCAGCT	GCTCCTGTTC	TGCCTCCTTA	TGGTGCCGGC	GGTGGTGGCT	ACAGCTCATC	3720
GGGCGGCAGC	AGTTCCGGTG	GCAGCAGCAA	GAAAAAGCAC	AGCGATCGTG	ACCGCGACAA	3780
GGAGAGCAAA	AAGAATAAGA	GCCAAGACTA	CGCGAAGTAC	AATGGCGCTG	GTGGCGGCAT	3840
CTTTAATCCC	CTTGGCGGTG	CTGGCGCCGC	ACCCAATATG	TCTGGAGGAA	TGGGCGCCCC	3900
CATGTCTACT	GCTGTACAC	CATCCATGCT	GTTGGCGCCC	ACCGGTGCAG	TACCACCCTC	3960
TGCCGCTGGG	CTGGCACCGC	CTCCCATGCC	CGTCTACAAC	AAGAAGTAGT	GGTAGCGGTC	4020
AGAGGGTTAT	TCTTAAGTCG	TACGTTTG	TATATGTATA	GAACCTCAGT	AAGTCCGATT	4080
GTAGTATAGT	TGTTAGGATT	GTTAGTGAGA	TGCATTATTG	ATTTAGTTA	AGCACATAGA	4140
TAAAACCCA	AATTGGAAGT	GAAACCGGAT	GCGCAGATCG	AAGAAGAATG	GAAGTAGATG	4200
TCGCGATGGG	GCTGGACGTA	AAAGCAGTAC	TCAAATCGCG	AAAACTTTG	TACAGCATTA	4260
ATTAGTTAT	AACTATAATA	AATAGCATAAC	ATATAAGCCC	AAAAAAAAAA	AAAAAAAAAA	4320
AAAAAAA						4328

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1097 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ser Leu Leu Ala Thr Pro Met Pro Gln Ala Ala Thr Ala Ser Ser
1 5 10 15

Ser Ser Ser Ala Ser Ala Ala Ser Ala Ser Gly Ile Pro Ile Thr
20 25 30

Ala Asn Asn Asn Leu Pro Phe Glu Lys Asp Lys Ile Trp Tyr Phe Ser
35 40 45

Asn Asp Gln Leu Ala Asn Leu Pro Ser Arg Arg Cys Gly Ile Lys Gly
50 55 60

Asp Asp Glu Leu Gln Tyr Arg Gln Met Thr Ala Tyr Leu Ile Gln Glu
65 70 75 80

Met Gly Gln Arg Leu Gln Val Ser Gln Leu Cys Ile Asn Thr Ala Ile
85 90 95

Val Tyr Met His Arg Phe Tyr Ala Phe His Ser Phe Thr His Phe His
100 105 110

Arg Asn Ser Met Ala Ser Ala Ser Leu Phe Leu Ala Ala Lys Val Glu
115 120 125

Glu Gln Pro Arg Lys Leu Glu His Val Ile Arg Ala Ala Asn Lys Cys
130 135 140

Leu Pro Pro Thr Thr Glu Gln Asn Tyr Ala Glu Leu Ala Gln Glu Leu
145 150 155 160

Val Phe Asn Glu Asn Val Leu Leu Gln Thr Leu Gly Phe Asp Val Ala
165 170 175

Ile Asp His Pro His Thr His Val Val Arg Thr Cys Gln Leu Val Lys
180 185 190

Ala Cys Lys Asp Leu Ala Gln Thr Ser Tyr Phe Leu Ala Ser Asn Ser
195 200 205

Leu His Leu Thr Ser Met Cys Leu Gln Tyr Arg Pro Thr Val Val Ala
210 215 220

Cys Phe Cys Ile Tyr Leu Ala Cys Lys Trp Ser Arg Trp Glu Ile Pro
225 230 235 240

Gln Ser Thr Glu Gly Lys His Trp Phe Tyr Tyr Val Asp Lys Thr Val
245 250 255

Ser Leu Asp Leu Leu Lys Gln Leu Thr Asp Glu Phe Ile Ala Ile Tyr
260 265 270

Glu Lys Ser Pro Ala Arg Leu Lys Ser Lys Leu Asn Ser Ile Lys Ala
 275 280 285
 Ile Ala Gln Gly Ala Ser Asn Arg Thr Ala Asn Ser Lys Asp Lys Pro
 290 295 300
 Lys Glu Asp Trp Lys Ile Thr Glu Met Met Lys Gly Tyr His Ser Asn
 305 310 315 320
 Ile Thr Thr Pro Pro Glu Leu Leu Asn Gly Asn Asp Ser Arg Asp Arg
 325 330 335
 Asp Arg Asp Arg Glu Arg Glu Arg Glu Arg Glu Arg Asp Pro Ser Ser
 340 345 350
 Leu Leu Pro Pro Pro Ala Met Val Pro Gln Gln Arg Arg Gln Asp Gly
 355 360 365
 Gly His Gln Arg Ser Ser Ser Val Ser Gly Val Pro Gly Ser Ser Ser
 370 375 380
 Ser Ser Ser Ser Ser His Lys Met Pro Asn Tyr Pro Gly Gly Met
 385 390 395 400
 Pro Pro Glu Ala His Pro Asp His Lys Ser Lys Gln Pro Gly Tyr Asn
 405 410 415
 Asn Arg Met Pro Ser Ser His Gln Arg Ser Ser Ser Ser Gly Leu Gly
 420 425 430
 Ser Ser Gly Ser Gly Ser Gln His Ser Ser Ser Ser Ser Ser Ser Ser
 435 440 445
 Ser Gln Gln Pro Gly Arg Pro Ser Met Pro Val Asp Tyr His Lys Ser
 450 455 460
 Ser Arg Gly Met Pro Pro Val Gly Val Gly Met Pro Pro His Gly Ser
 465 470 475 480
 His Lys Met Thr Ser Gly Ser Lys Pro Gln Gln Pro Gln Gln Pro
 485 490 495
 Val Pro His Pro Ser Ala Ser Asn Ser Ser Ala Ser Gly Met Ser Ser
 500 505 510
 Lys Asp Lys Ser Gln Ser Asn Lys Met Tyr Pro Asn Ala Pro Pro Pro
 515 520 525
 Tyr Ser Asn Ser Ala Pro Gln Asn Pro Leu Met Ser Arg Gly Gly Tyr
 530 535 540
 Pro Gly Ala Ser Asn Gly Ser Gln Pro Pro Pro Pro Ala Gly Tyr Gly
 545 550 555 560

Gly His Arg Ser Lys Ser Gly Ser Thr Val His Gly Met Pro His Phe
565 570 575

Glu Gln Gln Leu Pro Tyr Ser Gln Ser Gln Ser Tyr Gly His Met Gln
580 585 590

Gln Gln Pro Val Pro Gln Ser Gln Gln Gln Met Pro Pro Glu Ala
595 600 605

Ser Gln His Ser Leu Gln Ser Lys Asn Ser Leu Phe Ser Pro Glu Trp
610 615 620

Pro Asp Ile Lys Lys Glu Pro Met Ser Gln Ser Gln Pro Gln Leu Phe
625 630 635 640

Asn Gly Leu Leu Pro Pro Pro Ala Pro Pro Gly His Asp Tyr Lys Leu
645 650 655

Asn Ser His Pro Arg Asp Lys Glu Ser Pro Lys Lys Glu Arg Leu Thr
660 665 670

Pro Thr Lys Lys Asp Lys His Arg Pro Val Met Pro Pro Met Gly Ser
675 680 685

Gly Asn Ser Ser Ser Gly Ser Gly Ser Ser Lys Pro Met Leu Pro Pro
690 695 700

His Lys Lys Gln Ile Pro His Gly Gly Asp Leu Leu Thr Asn Pro Gly
705 710 715 720

Glu Ser Gly Ser Leu Lys Arg Pro Asn Glu Ile Ser Gly Ser Gln Tyr
725 730 735

Gly Leu Asn Lys Leu Asp Glu Ile Asp Asn Ser Asn Met Pro Arg Glu
740 745 750

Lys Leu Arg Lys Leu Asp Thr Thr Thr Gly Leu Pro Thr Tyr Pro Asn
755 760 765

Tyr Glu Glu Lys His Thr Pro Leu Asn Met Ser Asn Gly Ile Glu Thr
770 775 780

Thr Pro Asp Leu Val Arg Ser Leu Leu Lys Glu Ser Leu Cys Pro Ser
785 790 795 800

Asn Ala Ser Leu Leu Lys Pro Asp Ala Leu Thr Met Pro Gly Leu Lys
805 810 815

Pro Pro Ala Glu Leu Leu Glu Pro Met Pro Ala Pro Ala Thr Ile Lys
820 825 830

Lys Glu Gln Gly Ile Thr Pro Met Thr Ser Leu Ala Ser Gly Pro Ala
835 840 845

Pro Met Asp Leu Glu Val Pro Thr Lys Gln Ala Gly Glu Ile Lys Glu
 850 855 860

 Glu Ser Ser Ser Lys Ser Glu Lys Lys Lys Lys Lys Asp Lys His Lys
 865 870 875 880

 His Lys Glu Lys Asp Lys Ser Lys Asp Lys Thr Glu Lys Glu Glu Arg
 885 890 895

 Lys Lys His Lys Arg Asp Lys Gln Lys Asp Arg Ser Gly Ser Gly Gly
 900 905 910

 Ser Lys Asp Ser Ser Leu Pro Asn Glu Pro Leu Lys Met Val Ile Lys
 915 920 925

 Asn Pro Asn Gly Ser Leu Gln Ala Gly Ala Ser Ala Pro Ile Lys Leu
 930 935 940

 Lys Ile Ser Lys Asn Lys Val Glu Pro Asn Asn Tyr Ser Ala Ala Ala
 945 950 955 960

 Gly Leu Pro Gly Ala Ile Gly Tyr Gly Leu Pro Pro Thr Thr Ala Thr
 965 970 975

 Thr Thr Ser Ala Ser Ile Gly Ala Ala Ala Pro Val Leu Pro Pro Tyr
 980 985 990

 Gly Ala Gly Gly Gly Tyr Ser Ser Ser Gly Gly Ser Ser Ser Gly
 995 1000 1005

 Gly Ser Ser Lys Lys His Ser Asp Arg Asp Arg Asp Lys Glu Ser
 1010 1015 1020

 Lys Lys Asn Lys Ser Gln Asp Tyr Ala Lys Tyr Asn Gly Ala Gly Gly
 1025 1030 1035 1040

 Gly Ile Phe Asn Pro Leu Gly Gly Ala Ala Ala Pro Asn Met Ser
 1045 1050 1055

 Gly Gly Met Gly Ala Pro Met Ser Thr Ala Val Pro Pro Ser Met Leu
 1060 1065 1070

 Leu Ala Pro Thr Gly Ala Val Pro Pro Ser Ala Ala Gly Leu Ala Pro
 1075 1080 1085

 Pro Pro Met Pro Val Tyr Asn Lys Lys
 1090 1095

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1119 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..1116

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATG GCA AAG CAG TAC GAC TCG GTG GAG TGC CCT TTT TGT GAT GAA GTT	48
Met Ala Lys Gln Tyr Asp Ser Val Glu Cys Pro Phe Cys Asp Glu Val	
1 5 10 15	
TCC AAA TAC GAG AAG CTC GCC AAG ATC GGC CAA GGC ACC TTC GGG GAG	96
Ser Lys Tyr Glu Lys Leu Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu	
20 25 30	
GTG TTC AAG GCC AGG CAC CGC AAG ACC GGC CAG AAG GTG GCT CTG AAG	144
Val Phe Lys Ala Arg His Arg Lys Thr Gly Gln Lys Val Ala Leu Lys	
35 40 45	
AAG GTG CTG ATG GAA AAC GAG AAG GAG GGG TTC CCC ATT ACA GCC TTG	192
Lys Val Leu Met Glu Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu	
50 55 60	
CGG GAG ATC AAG ATC CTT CAG CTT CTA AAA CAC GAG AAT GTG GTC AAC	240
Arg Glu Ile Lys Ile Leu Gln Leu Lys His Glu Asn Val Val Asn	
65 70 75 80	
TTG ATT GAG ATT TGT CGA ACC AAA GCT TCC CCC TAT AAC CGC TGC AAG	288
Leu Ile Glu Ile Cys Arg Thr Lys Ala Ser Pro Tyr Asn Arg Cys Lys	
85 90 95	
GGT AGT ATA TAC CTG GTG TTC GAC TTC TGC GAG CAT GAC CTT GCT GGG	336
Gly Ser Ile Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly	
100 105 110	
CTG TTG AGC AAT GTT TTG GTC AAG TTC ACG CTG TCT GAG ATC AAG AGG	384
Leu Leu Ser Asn Val Leu Val Lys Phe Thr Leu Ser Glu Ile Lys Arg	
115 120 125	
GTG ATG CAG ATG CTG CTT AAC GGC CTC TAC TAC ATC CAC AGA AAC AAG	432
Val Met Gln Met Leu Leu Asn Gly Leu Tyr Tyr Ile His Arg Asn Lys	
130 135 140	
ATC CTG CAT AGG GAC ATG AAG GCT GCT AAT GTG CTT ATC ACT CGT GAT	480
Ile Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Arg Asp	
145 150 155 160	
GGG GTC CTG AAG CTG GCA GAC TTT GGG CTG GCC CGG GCC TTC AGC CTG	528
Gly Val Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Leu	
165 170 175	

0805210808-102592

GCC AAG AAC AGC CAG CCC AAC CGC TAC ACC AAC CGT GTG GTG ACA CTC Ala Lys Asn Ser Gln Pro Asn Arg Tyr Thr Asn Arg Val Val Thr Leu 180 185 190	576
TGG TAC CGG CCC CCG GAG CTG TTG CTC GGG GAG CGG GAC TAC GGC CCC Trp Tyr Arg Pro Pro Glu Leu Leu Gly Glu Arg Asp Tyr Gly Pro 195 200 205	624
CCC ATT GAC CTG TGG GGT GCT GGG TGC ATC ATG GCA GAG ATG TGG ACC Pro Ile Asp Leu Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr 210 215 220	672
CGC AGC CCC ATC ATG CAG GGC AAC ACG GAG CAG CAC CAA CTC GCC CTC Arg Ser Pro Ile Met Gln Gly Asn Thr Glu Gln His Gln Leu Ala Leu 225 230 235 240	720
ATC AGT CAG CTC TGC GGC TCC ATC ACC CCT GAG GTG TGG CCA AAC GTG Ile Ser Gln Leu Cys Gly Ser Ile Thr Pro Glu Val Trp Pro Asn Val 245 250 255	768
GAC AAC TAT GAG CTG TAC GAA AAG CTG GAG CTG GTC AAG GGC CAG AAG Asp Asn Tyr Glu Leu Tyr Glu Lys Leu Glu Leu Val Lys Gly Gln Lys 260 265 270	816
CGG AAG GTG AAG GAC AGG CTG AAG GCC TAT GTG CGT GAC CCA TAC GCA Arg Lys Val Lys Asp Arg Leu Lys Ala Tyr Val Arg Asp Pro Tyr Ala 275 280 285	864
CTG GAC CTC ATC GAC AAG CTG CTG GTG CTG GAC CCT GCC CAG CGC ATC Leu Asp Leu Ile Asp Lys Leu Leu Val Leu Asp Pro Ala Gln Arg Ile 290 295 300	912
GAC AGC GAT GAC GCC CTC AAC CAC GAC TTC TTC TGG TCC GAC CCC ATG Asp Ser Asp Asp Ala Leu Asn His Asp Phe Phe Trp Ser Asp Pro Met 305 310 315 320	960
CCC TCC GAC CTC AAG GGC ATG CTC TCC ACC CAC CTG ACG TCC ATG TTC Pro Ser Asp Leu Lys Gly Met Leu Ser Thr His Leu Thr Ser Met Phe 325 330 335	1008
GAG TAC TTG GCA CCA CCG CGC CGG AAG GGC AGC CAG ATC ACC CAG CAG Glu Tyr Leu Ala Pro Pro Arg Arg Lys Gly Ser Gln Ile Thr Gln Gln 340 345 350	1056
TCC ACC AAC CAG AGT CGC AAT CCC GCC ACC ACC AAC CAG ACG GAG TTT Ser Thr Asn Gln Ser Arg Asn Pro Ala Thr Thr Asn Gln Thr Glu Phe 355 360 365	1104
GAG CGC GTC TTC TGA Glu Arg Val Phe 370	1119

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 372 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Lys Gln Tyr Asp Ser Val Glu Cys Pro Phe Cys Asp Glu Val
1 5 10 15

Ser Lys Tyr Glu Lys Leu Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu
20 25 30

Val Phe Lys Ala Arg His Arg Lys Thr Gly Gln Lys Val Ala Leu Lys
35 40 45

Lys Val Leu Met Glu Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu
50 55 60

Arg Glu Ile Lys Ile Leu Gln Leu Leu Lys His Glu Asn Val Val Asn
65 70 75 80

Leu Ile Glu Ile Cys Arg Thr Lys Ala Ser Pro Tyr Asn Arg Cys Lys
85 90 95

Gly Ser Ile Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly
100 105 110

Leu Leu Ser Asn Val Leu Val Lys Phe Thr Leu Ser Glu Ile Lys Arg
115 120 125

Val Met Gln Met Leu Leu Asn Gly Leu Tyr Tyr Ile His Arg Asn Lys
130 135 140

Ile Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Arg Asp
145 150 155 160

Gly Val Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Leu
165 170 175

Ala Lys Asn Ser Gln Pro Asn Arg Tyr Thr Asn Arg Val Val Thr Leu
180 185 190

Trp Tyr Arg Pro Pro Glu Leu Leu Gly Glu Arg Asp Tyr Gly Pro
195 200 205

Pro Ile Asp Leu Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr
210 215 220

Arg Ser Pro Ile Met Gln Gly Asn Thr Glu Gln His Gln Leu Ala Leu
225 230 235 240

Ile Ser Gln Leu Cys Gly Ser Ile Thr Pro Glu Val Trp Pro Asn Val
245 250 255

Asp Asn Tyr Glu Leu Tyr Glu Lys Leu Glu Leu Val Lys Gly Gln Lys
260 265 270

Arg Lys Val Lys Asp Arg Leu Lys Ala Tyr Val Arg Asp Pro Tyr Ala
275 280 285

Leu Asp Leu Ile Asp Lys Leu Leu Val Leu Asp Pro Ala Gln Arg Ile
290 295 300

Asp Ser Asp Asp Ala Leu Asn His Asp Phe Phe Trp Ser Asp Pro Met
305 310 315 320

Pro Ser Asp Leu Lys Gly Met Leu Ser Thr His Leu Thr Ser Met Phe
325 330 335

Glu Tyr Leu Ala Pro Pro Arg Arg Lys Gly Ser Gln Ile Thr Gln Gln
340 345 350

Ser Thr Asn Gln Ser Arg Asn Pro Ala Thr Thr Asn Gln Thr Glu Phe
355 360 365

Glu Arg Val Phe
370

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACGAATTCCA CACAATCCAA AGATC

25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CAGAATTCCCT ATTGCCGATC CCCAGA

26

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: one-of(8, 14)
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A or C or G or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "Y = C or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: one-of(17, 20)
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGAATTCNAT GYTNCARCAR CC

22

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: one-of(13, 16, 19, 22, 25)
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AACTGCAGTC CARAARAART CRTGRTT

27

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGTCAAGGAT CAAACCGGCT GTGAT

25

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGAATTCCAA GAAACGCATC GATGC

25

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGACCTGCCA AATCGTGT

18

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGAAGGTGGA TCTGTAACCA TTCGT

25

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGAATTCA GAGA TCTCGATCAG ATTCA

25

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TTACTACTCG AGCTTACCAAAA CCCGGTC

27

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TAAGCAAGCT TCTATGGCGC ACATGTCC

28

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TTACTACTCG AGCTTACCAAA CCCGGTC

27

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: one-of(13, 16, 22)
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "Y = C or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 17
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "W = A or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "S = C or G"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 19
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A or C or G or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGAATTCTGG TAYTTYWSNA AYGA

24

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "Y = C or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 14

(D) OTHER INFORMATION: /mod_base= OTHER
/note= "R = A or G"

(ix) FEATURE:

(A) NAME/KEY: modified_base
(B) LOCATION: one-of(17, 20)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = A or C or G or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGGGATCCTG YTCRAANGGN GGCAT

25

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: modified_base
(B) LOCATION: one-of(11, 14, 20)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = A or C or G or T"

(ix) FEATURE:

(A) NAME/KEY: modified_base
(B) LOCATION: 23
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGGGATCCAA NGGNNGGCATN CCRT

24

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ATCACGACAC CACCAGAGCT GTTA

24

03951188-1021597
(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGAATTCAAGA TCGTGAACGG GA

22

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CGAATTCAAGG CGCTAGCAAT G

21

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GAAAGGCGTA GAACCGA

17

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCTGACCCAT TTCCTGTATC AGATAG

26

2653701 2653701 2653701

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAATTCTTC TGCTTGGCGA AT

22

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGGAATTCGA GGTTCTATAC ATAT

24

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CTGTGTGAAT GGAATCTGTG ATGTG

25

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TATCCCGGGT CATATGAGTC TCCTAGCC

28

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met Leu Gln Gln Pro Ser Gly Ser Thr Pro Ser Asn Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Ala Asp Thr Ala Leu Asn His Asp Phe Phe Trp Thr Asp Pro Met Pro
1 5 10 15

Ser

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met Leu Gln Gln Pro
1 5

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Asn His Asp Phe Phe Trp Thr
1 5

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Ser Pro Glu Trp Pro Asp Ile
1 5

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Trp Tyr Phe Ser Asn Asp Gln Leu Ala Asn Ser Pro Ser Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Thr Val His Gly Met Pro Pro Phe Glu Gln Gln Leu Pro Tyr
1 5 10

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Trp Tyr Phe Ser Asn Asp
1 5

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Met Pro Pro Phe Glu Gln
1 5

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

His Gly Met Pro Pro Phe
1 5

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCAGGATCCA GAATTCCATA TGGCAAAGCA GTACGACTCG G

41

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CAGTACTCGA GTTATCAGAA GACGCGCTCA AAC

33

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4528 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGGGGGGGGG	GGGTGAATGA	AGGAGCGGGC	GGAGGGAGGAA	TTGTCATGGC	GTCGGGCCGT	60
GGAGCTTCTT	CTCGCTGGTT	CTTTACTCGG	GAACAGCTGG	AGAACACGCC	GAGCCGCCGC	120
TGCGGAGTGG	AGCGGATAA	AGAGCTCTCG	TGCCGCCAGC	AGGCGGCCAA	CCTCATCCAG	180
GAGATGGGAC	AGCGTCTCAA	TGTCTCTCAG	CTTACAATAA	ACACTGCGAT	TGTTTATATG	240
CACAGGTTTT	ATATGCACCA	TTCTTCACC	AAATTCAACA	AAAATATAAT	ATCGTCTACT	300
GCATTATTTT	TGGCTGCAAA	AGTGGAAAGAA	CAGGCTCGAA	AACTTGAACA	TGTTATCAAA	360
GTAGCACATG	CTTGTCTTCA	TCCTCTAGAG	CCACTGCTGG	ATACTAAATG	TGATGCTTAC	420
CTTCAACAGA	CTCAAGAACT	GGTTATACTT	GAAACCATAA	TGCTACAAAC	TCTAGGTTTT	480
GAGATCACCA	TTGAACACCCC	ACACACAGAT	GTGGTGAAAT	GTACCCAGTT	AGTAAGAGCA	540
AGCAAGGATT	TGGCACAGAC	ATCCTATTTC	ATGGCTACCA	ACAGTCTGCA	TCTTACAACC	600
TTCTGTCTTC	AGTACAAACC	AACAGTGATA	GCATGTGTAT	GCATTCAATT	GGCTTGCAAA	660
TGGTCCAATT	GGGAGATCCC	TGTATCAACT	GATGGAAAGC	ATTGGTGGGA	ATATGTGGAT	720
CCTACAGTTA	CTCTAGAATT	ATTAGATGAG	CTAACACATG	AGTTTCTACA	AATATTGGAG	780
AAAACGCCTA	ATAGGTTGAA	GAAGATTCGA	AACTGGAGGG	CTAACAGGC	AGCTAGGAAA	840
CCAAAAGTAG	ATGGACAGGT	ATCAGAGACA	CCACTTCTTG	GTTCATCTTT	GGTCCAGAAT	900
TCCATTTAG	TAGATAGTGT	CACTGGTGTG	CCTACAAACC	CAAGTTTCA	GAAACCATCT	960
ACATCAGCAT	TCCCTGCGCC	AGTACCTCTA	AATTCAAGGAA	ATATTCTGT	TCAAGACAGC	1020
CATACATCTG	ATAATTTGTC	AATGCTAGCA	ACAGGAATGC	CAAGTACTTC	ATACGGTTA	1080

00000000000000000000000000000000

TCATCACACC	AGGAATGGCC	TCAACATCAA	GACTCAGCAA	GGACAGAACAA	GCTATATTCA	1140
CAGAAACAGG	AGACATCTT	GTCTGGTAGC	CAGTACAACA	TCAACTTCCA	GCAGGGACCT	1200
TCTATATCAC	TGCATTCAGG	ATTACATCAC	AGACCTGACA	AAATTCAGA	TCATTCTCT	1260
GTAAAGCAAG	AATATACTCA	TAAAGCAGGG	AGCAGTAAAC	ACCATGGGCC	AATTTCCACT	1320
ACTCCAGGAA	TAATTCCCTCA	GAAAATGTCT	TTAGATAAAAT	ATAGAGAAAAA	GCGTAAACTA	1380
GAAACTCTTG	ATCTCGATGT	AAGGGATCAT	TATATAGCTG	CCCAGGTAGA	ACAGCAGCAC	1440
AAACAAGGGC	AGTCACAGGC	AGCCAGCAGC	AGTTCTGTTA	CTTCTCCCAT	AAAAATGAAA	1500
ATACCTATCG	CAAATACTGA	AAAATACATG	GCAGATAAAA	AGGAAAAGAG	TGGGTCACTG	1560
AAATTACGGA	TTCCAATACC	ACCCACTGAT	AAAAGCGCCA	GTAAAGAAGA	ACTGAAAATG	1620
AAAATAAAAG	TTTCTTCTTC	AGAAAGACAC	AGCTCTCTG	ATGAAGGCAG	TGGGAAAAGC	1680
AAACATTCAA	GCCCACATAT	TAGCAGAGAC	CATAAGGAGA	AGCACAAGGA	GCATCCTTCA	1740
AGCCGCCACC	ACACCAGCAG	CCACAAGCAT	TCCCACTCGC	ATAGTGGCAG	CAGCAGCGGT	1800
GGCAGTAAAC	ACAGTGCCGA	CGGAATACCA	CCCACGTTC	TGAGGAGTCC	TGTTGGCCTG	1860
AGCAGTGATG	GCATTTCCCTC	TAGCTCCAGC	TCTTCAAGGA	AGAGGCTGCA	TGTCAATGAT	1920
GCATCTCACA	ACCACCACTC	CAAATGAGC	AAAAGTTCCA	AAAGTTCAGG	TGGGCTACGG	1980
ACATCTCAGC	ACCTCGTGAA	ACTGGACAAG	AAGCCAGTGG	AGACCAACGG	TCCTGATGCC	2040
AATCACGAGT	ACAGTACAAG	CAGCCAGCAT	ATGGACTACA	AAGACACATT	CGACATGCTG	2100
GAECTCACTGT	TAAGTGCCCA	AGGAATGAAC	ATGTAATAAT	TTGTTTAGGT	CAATTTTCC	2160
TTTACTTTTT	TAATTAAAAA	ATTGTTAGAA	TGGAAAATT	CCTTCTGATC	TAGCAGTGGT	2220
AACCCCTGCT	GTTGCTGCCA	CTGCTTCAAT	ATTTGTAAGT	GCTACTTTAT	TCTTCATTCT	2280
GAAAAGAAGA	GATTATAGTA	AACAAGTCTT	TATCTCCACA	TATGATAGTG	TTATAAATAC	2340
TGTAAAGGCA	TGGAAGGTGC	AAAACTCAGT	ATTTCTACAA	TTGCAGCTAA	GAACATTAGG	2400
ATGAATGGCT	GGCTGCTTCT	AGGAATATAA	GATGCCTCAA	GCATTCAATTA	TTTATGATTT	2460
GAATACTGTA	GCTATTTTTT	GTTGCTTGGC	TTTGAAATGA	GTGTAAATTG	TTTTCTTTG	2520
TGTATTTATA	CTTGTATGTA	TGATTTGCAT	GTTCATGAA	TAAAGGGATA	AAACAGTATA	2580
CTGACAACTG	TTTACAAGAA	AGTGGAGAAA	ATGTAACACAA	TTTTGTATGT	TTAGATATTA	2640
CCGTAATAC	TCAGGATTGG	AGCTGCTTGT	AAGTATAACAA	ATATACAGAA	TACTTTATTT	2700

TATCTTGTCA GAGTCCATC ACTATCTAAA ACAAAGGTGC AATTTTTAT GTTAACCTTA	2760
AATCTAGCCC TTACTGGAAG CCACTGATAG GGACATTAC TACAGATGT GTGCAGTGCA	2820
GCAGATGGTC ATATAACACT GTGAGGCCTA GAATTTGCC TTCAGAGGTT CTGACCAGAT	2880
TGGCTGCTGA AATAGCCCCT AACTTCTGA AGGCTTGAAG AGGAAAAAAT AAAGTTTACA	2940
TACTCTTGAT GGAAGTGCAT TTAAATGTT GTTGGCTTGT TGCAGTTCTA TGAAACAGAG	3000
CTGTTAATAA TGGTTATGTG GATTACTGTG ATTTGAAAAC TAAATTACACA ATAACCTTACC	3060
TAGTAGAGAT TTAGTGAGTT GTTCCCTTA AAGAATTAA CACTACATAT TTTAATAGTA	3120
AACAGGGTCA CTTCCCTTA GCATTCAGAA TGACACCATA TTCTTAAATA TACTCCTTCC	3180
CTGAAGCGTG TTTGTGTGTG ATGCCATATT TCTTTTCAG GTAAATGTAG TCTTCCTTAT	3240
AAAAATGAAA TTAAACCTAT GCTCTCAATT CTTTATATT CTAACAATAA ATAAAAAAAGA	3300
AAAGATTACT GACTGTGCAT TGTACCTGTA TTTATAGTT ATGGTTATCA GAAGCTCTGT	3360
AAGAAAGAAA AGTCAGCTC CCAGGCAAAC CAGTAGTGGA GGTTTACAT TTGTTGCAC	3420
ATCTCAGTAT ATTTCTGTTG AGGTAAAGTT TGCACAGTCA TCTGACTTCT GATCAAGCAT	3480
TAGATTTAA CTTGTTAGA TTTGTCTTA AACACCAGTA ATATGGCTCT TGTTATCAG	3540
CTAATCTTGA ATTTATTCTG TGGTAAATCT TTTGAGTTGC TGAGTATATT TGAGATTGAT	3600
TGGATTCAAC CTCTGTTGA ACTGAAAATCT TAATTTTTC TCTGTATTT TGTTACAAAG	3660
CCACTGATAC GTGCACAATT GTAATTAAGT ATGTTGCAGT TGTAATATT AGAGTTAAT	3720
CTCATGCTCT ACCTTTATTT AGCAATTACC TAATTTGCCA GTAGCTTAT AATTTTAAA	3780
GATAATTGTT CATTATTTG TCAATGTTAT TTGAACCTGG GGTACTTAGG AGCCTCTTG	3840
TAGGGACTGT GCCTAGGTAG CATGCTCTAA CATTGTTCT GGTCTTGCAT AACCTCAGTA	3900
TCTTGTCAT TATATGTAAC TTTGTTGCTC TGTATGGCAT AATATTGTAT CCATAAACAT	3960
GGTAATTTG ATACAGTTAT ACTTTACAG TGGTACATAA TCCAAGGACT AGTATAGAAT	4020
TAAGCTGAGT GCAAGATGAG GGAGGGAAAGG GCTTCTTGG TAATTTAGAT GTGAAACCTC	4080
TACAGAGCTA TCATGAAAA ACTACATGAG GTGGTTGTGC TACTGTATAA TTGGGGGTGA	4140
TAATACCAGG AATTTAATA AGATTTGTA AAGAATATCC AGAAAAGTAG TGAACATTATT	4200
TTCAGTAGGC ATAGAAAACA ATGTGAATAT TTAAGGTCTG TGACTATAGT TAAACTTCAC	4260
TAAGAATTG CAGAATTGTT TTGAGATGTG TGAATAAAGG TAATTTATT GAATCTTCAT	4320

TGGTGCTAAT GTTGGACAGT TAAAAAGATA GCTAGTGTAT ATTGTTATGG GTCAGTACTT	4380
ATTAGTACTT CCAAAATTGA ATTTGAAATG CTATGTATTC ACTTTCACT CTGTAAATGT	4440
AATTCTTAC AATGACTTTA TTTATTAAAG GGCAGCCAGT TGTCATTGT AAAAAAAAAA	4500
AAAAAAAAAA AAAGCGGCCG CTGAATTC	4528

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2091 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

ATGGCGTCGG GCCGTGGAGC TTCTTCTCGC TGGTTCTTA CTCGGAAACA GCTGGAGAAC	60
ACGCCGAGCC GCCGCTGCGG AGTGGAGGCG GATAAAGAGC TCTCGTGCCG CCAGCAGGCG	120
GCCAACCTCA TCCAGGGAGAT GGGACAGCGT CTCAATGTCT CTCAGCTTAC AATAAACACT	180
GCGATTGTTT ATATGCACAG GTTTTATATG CACCATTCTT TCACCAAATT CAACAAAAAT	240
ATAATATCGT CTACTGCATT ATTTTGCT GCAAAAGTGG AAGAACAGGC TCGAAAACCTT	300
GAACATGTTA TCAAAGTAGC ACATGCTTGT CTTCATCCTC TAGAGCCACT GCTGGATACT	360
AAATGTGATG CTTACCTTCA ACAGACTCAA GAACTGGTTA TACTTGAAAC CATAATGCTA	420
CAAACCTCTAG GTTTGAGAT CACCATTGAA CACCCACACA CAGATGTGGT GAAATGTACC	480
CAGTTAGTAA GAGCAAGCAA GGATTTGGCA CAGACATCCT ATTTCATGGC TACCAACAGT	540
CTGCATCTTA CAACCTCTG TCTTCAGTAC AAACCAACAG TGATAGCATG TGTATGCATT	600
CATTTGGCTT GCAAATGGTC CAATTGGAG ATCCCTGTAT CAACTGATGG AAAGCATTGG	660
TGGGAATATG TGGATCCTAC AGTTACTCTA GAATTATTAG ATGAGCTAAC ACATGAGTTT	720
CTACAAATAT TGGAGAAAAC GCCTAATAGG TTGAAGAAGA TTGAAACTG GAGGGCTAAT	780
CAGGCAGCTA GGAAACCAAAGTAGAGATGGA CAGGTATCAG AGACACCACT TCTTGGTTCA	840
TCTTTGGTCC AGAATTCCAT TTTAGTAGAT AGTGTCACTG GTGTGCCTAC AAACCCAAGT	900
TTTCAGAAAC CATCTACATC AGCATTCCCT GCGCCAGTAC CTCTAAATTC AGGAAATATT	960
TCTGTTCAAG ACAGCCATAC ATCTGATAAT TTGTCAATGC TAGCAACAGG AATGCCAAGT	1020
ACTTCATACG GTTTATCATC ACACCAGGAA TGGCCTCAAC ATCAAGACTC AGCAAGGACA	1080

GAACAGCTAT ATTACAGAA ACAGGAGACA TCTTGCTG GTAGCCAGTA CAACATCAAC 1140
 TTCCAGCAGG GACCTTCTAT ATCACTGCAT TCAGGATTAC ATCACAGACC TGACAAAATT 1200
 TCAGATCATT CTTCTGTTAA GCAAGAATAT ACTCATAAAG CAGGGAGCAG TAAACACCAT 1260
 GGGCCAATT CCACTAACCC AGGAATAATT CCTCAGAAAA TGTCTTTAGA TAAATATAGA 1320
 GAAAAGCGTA AACTAGAAAC TCTTGATCTC GATGTAAGGG ATCATTATAT AGCTGCCAG 1380
 GTAGAACAGC AGCACAAACA AGGGCAGTCA CAGGCAGCCA GCAGCAGTTC TGTTACTTCT 1440
 CCCATTAAAA TGAAAATACC TATCGCAAAT ACTGAAAAT ACATGGCAGA TAAAAAGGAA 1500
 AAGAGTGGGT CACTGAAATT ACGGATTCCA ATACCACCA CTGATAAAAG CGCCAGTAAA 1560
 GAAGAACTGA AAATGAAAAT AAAAGTTCT TCTTCAGAAA GACACAGCTC TTCTGATGAA 1620
 GGCAGTGGGA AAAGCAAACA TTCAAGCCA CATATTAGCA GAGACCATAA GGAGAACAC 1680
 AAGGAGCATC CTTCAAGCCG CCACCACACC AGCAGCCACA AGCATTCCCA CTCGCATAGT 1740
 GGCAGCAGCA GCGGTGGCAG TAAACACAGT GCCGACGGAA TACCACCCAC TGTTCTGAGG 1800
 AGTCCTGTTG GCCTGAGCAG TGATGGCATT TCCTCTAGCT CCAGCTCTTC AAGGAAGAGG 1860
 CTGCATGTCA ATGATGCATC TCACAACCAC CACTCCAAA TGAGCAAAG TTCCAAAAGT 1920
 TCAGGTGGGC TACGGACATC TCAGCACCTC GTGAAACTGG ACAAGAAGCC AGTGGAGACC 1980
 AACGGTCCTG ATGCCAATCA CGAGTACAGT ACAAGCAGCC AGCATATGGA CTACAAAGAC 2040
 ACATTGACA TGCTGGACTC ACTGTTAAGT GCCCAAGGAA TGAACATGTA A 2091

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 696 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Met Ala Ser Gly Arg Gly Ala Ser Ser Arg Trp Phe Phe Thr Arg Glu
 1 5 10 15

Gln Leu Glu Asn Thr Pro Ser Arg Arg Cys Gly Val Glu Ala Asp Lys
 20 25 30

Glu Leu Ser Cys Arg Gln Gln Ala Ala Asn Leu Ile Gln Glu Met Gly
 35 40 45

Gln Arg Leu Asn Val Ser Gln Leu Thr Ile Asn Thr Ala Ile Val Tyr
 50 55 60
 Met His Arg Phe Tyr Met His His Ser Phe Thr Lys Phe Asn Lys Asn
 65 70 75 80
 Ile Ile Ser Ser Thr Ala Leu Phe Leu Ala Ala Lys Val Glu Glu Gln
 85 90 95
 Ala Arg Lys Leu Glu His Val Ile Lys Val Ala His Ala Cys Leu His
 100 105 110
 Pro Leu Glu Pro Leu Leu Asp Thr Lys Cys Asp Ala Tyr Leu Gln Gln
 115 120 125
 Thr Gln Glu Leu Val Ile Leu Glu Thr Ile Met Leu Gln Thr Leu Gly
 130 135 140
 Phe Glu Ile Thr Ile Glu His Pro His Thr Asp Val Val Lys Cys Thr
 145 150 155 160
 Gln Leu Val Arg Ala Ser Lys Asp Leu Ala Gln Thr Ser Tyr Phe Met
 165 170 175
 Ala Thr Asn Ser Leu His Leu Thr Thr Phe Cys Leu Gln Tyr Lys Pro
 180 185 190
 Thr Val Ile Ala Cys Val Cys Ile His Leu Ala Cys Lys Trp Ser Asn
 195 200 205
 Trp Glu Ile Pro Val Ser Thr Asp Gly Lys His Trp Trp Glu Tyr Val
 210 215 220
 Asp Pro Thr Val Thr Leu Glu Leu Leu Asp Glu Leu Thr His Glu Phe
 225 230 235 240
 Leu Gln Ile Leu Glu Lys Thr Pro Asn Arg Leu Lys Lys Ile Arg Asn
 245 250 255
 Trp Arg Ala Asn Gln Ala Ala Arg Lys Pro Lys Val Asp Gly Gln Val
 260 265 270
 Ser Glu Thr Pro Leu Leu Gly Ser Ser Leu Val Gln Asn Ser Ile Leu
 275 280 285
 Val Asp Ser Val Thr Gly Val Pro Thr Asn Pro Ser Phe Gln Lys Pro
 290 295 300
 Ser Thr Ser Ala Phe Pro Ala Pro Val Pro Leu Asn Ser Gly Asn Ile
 305 310 315 320
 Ser Val Gln Asp Ser His Thr Ser Asp Asn Leu Ser Met Leu Ala Thr
 325 330 335

Gly Met Pro Ser Thr Ser Tyr Gly Leu Ser Ser His Gln Glu Trp Pro
340 345 350

Gln His Gln Asp Ser Ala Arg Thr Glu Gln Leu Tyr Ser Gln Lys Gln
355 360 365

Glu Thr Ser Leu Ser Gly Ser Gln Tyr Asn Ile Asn Phe Gln Gln Gly
370 375 380

Pro Ser Ile Ser Leu His Ser Gly Leu His His Arg Pro Asp Lys Ile
385 390 395 400

Ser Asp His Ser Ser Val Lys Gln Glu Tyr Thr His Lys Ala Gly Ser
405 410 415 /

Ser Lys His His Gly Pro Ile Ser Thr Thr Pro Gly Ile Ile Pro Gln
420 425 430

Lys Met Ser Leu Asp Lys Tyr Arg Glu Lys Arg Lys Leu Glu Thr Leu
435 440 445

Asp Leu Asp Val Arg Asp His Tyr Ile Ala Ala Gln Val Glu Gln Gln
450 455 460

His Lys Gln Gly Gln Ser Gln Ala Ala Ser Ser Ser Val Thr Ser
465 470 475 480

Pro Ile Lys Met Lys Ile Pro Ile Ala Asn Thr Glu Lys Tyr Met Ala
485 490 495

Asp Lys Lys Glu Lys Ser Gly Ser Leu Lys Leu Arg Ile Pro Ile Pro
500 505 510

Pro Thr Asp Lys Ser Ala Ser Lys Glu Glu Leu Lys Met Lys Ile Lys
515 520 525

Val Ser Ser Ser Glu Arg His Ser Ser Ser Asp Glu Gly Ser Gly Lys
530 535 540

Ser Lys His Ser Ser Pro His Ile Ser Arg Asp His Lys Glu Lys His
545 550 555 560

Lys Glu His Pro Ser Ser Arg His His Thr Ser Ser His Lys His Ser
565 570 575

His Ser His Ser Gly Ser Ser Ser Gly Gly Ser Lys His Ser Ala Asp
580 585 590

Gly Ile Pro Pro Thr Val Leu Arg Ser Pro Val Gly Leu Ser Ser Asp
595 600 605

Gly Ile Ser Ser Ser Ser Ser Ser Arg Lys Arg Leu His Val Asn
610 615 620

Asp Ala Ser His Asn His His Ser Lys Met Ser Lys Ser Ser Lys Ser
 625 630 635 640
 Ser Gly Gly Leu Arg Thr Ser Gln His Leu Val Lys Leu Asp Lys Lys
 645 650 655
 Pro Val Glu Thr Asn Gly Pro Asp Ala Asn His Glu Tyr Ser Thr Ser
 660 665 670
 Ser Gln His Met Asp Tyr Lys Asp Thr Phe Asp Met Leu Asp Ser Leu
 675 680 685
 Leu Ser Ala Gln Gly Met Asn Met
 690 695

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2190 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATGGCGTCGG	GCCGTGGAGC	TTCTTCTCGC	TGGTTCTTTA	CTCGGAAACA	GCTGGAGAAC	60
ACGCCGAGCC	GCCGCTGCGG	AGTGGAGGCG	GATAAAGAGC	TCTCGTGCCG	CCAGCAGGCG	120
GCCAACCTCA	TCCAGGAGAT	GGGACAGCGT	CTCAATGTCT	CTCAGCTTAC	AATAAACACT	180
GCGATTGTTT	ATATGCACAG	GTGTTATATG	CACCATTCTT	TCACCAAATT	CAACAAAAAT	240
ATAATATCGT	CTACTGCATT	ATTTTGCGCT	GCAAAAGTGG	AAGAACAGGC	TCGAAAACCTT	300
GAACATGTTA	TCAAAGTAGC	ACATGCTTGT	CTTCATCCTC	TAGAGCCACT	GCTGGATACT	360
AAATGTGATG	CTTACCTTCA	ACAGACTCAA	GAACTGGTTA	TACTTGAAAC	CATAATGCTA	420
CAAACCTCTAG	GTGTTGAGAT	CACCATTGAA	CACCCACACA	CAGATGTGGT	GAAATGTACC	480
CAGTTAGTAA	GAGCAAGCAA	GGATTTGGCA	CAGACATCCT	ATTTCATGGC	TACCAACAGT	540
CTGCATCTTA	CAACCTTCTG	TCTTCAGTAC	AAACCAACAG	TGATAGCATG	TGTATGCATT	600
CATTTGGCTT	GCAAATGGTG	CAATTGGGAG	ATCCCTGTAT	CAACTGATGG	AAAGCATTGG	660
TGGGAATATG	TGGATCCTAC	AGTTACTCTA	GAATTATTAG	ATGAGCTAAC	ACATGAGTTT	720
CTACAAATAT	TGGAGAAAAC	GCCTAATAGG	TTGAAGAAGA	TTCGAAACTG	GAGGGCTAAT	780
CAGGCAGCTA	GGAAACCAAA	AGTAGATGGA	CAGGTATCAG	AGACACCACT	TCTTGGTTCA	840

D
0
9
8
7
6
5
4
3
2
1
0

TCTTTGGTCC	AGAATTCCAT	TTTAGTAGAT	AGTGTCACTG	GTGTGCCTAC	AAACCCAAGT	900
TTTCAGAAAC	CATCTACATC	AGCATTCCCT	GCGCCAGTAC	CTCTAAATTC	AGGAAATATT	960
TCTGTTCAAG	ACAGCCATAC	ATCTGATAAT	TTGTCAATGC	TAGAACAGG	AATGCCAAGT	1020
ACTTCATACG	GTTTATCATC	ACACCAGGAA	TGGCCTAAC	ATCAAGACTC	AGCAAGGACA	1080
GAACAGCTAT	ATTCACAGAA	ACAGGAGACA	TCTTTGTCTG	GTAGCCAGTA	CAACATCAAC	1140
TTCCAGCAGG	GACCTTCTAT	ATCACTGCAT	TCAGGATTAC	ATCACAGACC	TGACAAAATT	1200
TCAGATCATT	CTTCTGTTAA	GCAGGAATAT	ACTCATAAAG	CAGGGAGCAG	TAAACACCAT	1260
GGGCCAATT	CCACTACTCC	AGGAATAATT	CCTCAGAAAA	TGTCTTAGA	TAAATATAGA	1320
GAAAAGCGTA	AACTAGAAAC	TCTTGATCTC	GATGTAAGGG	ATCATTATAT	AGCTGCCAG	1380
GTAGAACAGC	AGCACAAACA	AGGGCAGTCA	CAGGCAGCCA	GCAGCAGTTC	TGTTACTTCT	1440
CCCATTAAAA	TGAAAATACC	TATCGCAAAT	ACTGAAAAAT	ACATGGCAGA	TAAAAAGGAA	1500
AAGAGTGGGT	CACTGAAATT	ACGGATTCCA	ATACCACCA	CTGATAAAAG	CGCCAGTAAA	1560
GAAGAACTGA	AAATGAAAAT	AAAAGTTCT	TCTTCAGAAA	GACACAGCTC	TTCTGATGAA	1620
GGCAGTGGGA	AAAGCAAACA	TTCAAGCCC	CATATTAGCA	GAGACCATAA	GGAGAACAC	1680
AAGGAGCATC	CTTCAAGCCG	CCACCACACC	AGCAGCCACA	AGCATTCCCA	CTCGCATAGT	1740
GGCAGCAGCA	GCGGTGGCAG	TAAACACAGT	GCCGACGGAA	TACCACCCAC	TGTTCTGAGG	1800
AGTCCTGTTG	GCCTGAGCAG	TGATGGCATT	TCCTCTAGCT	CCAGCTCTTC	AAGGAAGAGG	1860
CTGCATGTCA	ATGATGCATC	TCACAACCAC	CACTCCAAA	TGAGCAAAG	TTCCAAAAGT	1920
TCAGGTAGTT	CATCTAGTTC	TTCCTCCTCT	GTAAAGCAGT	ATATATCCTC	TCACAACCTCT	1980
GTTTTTAAC	ATCCCTTACC	CCTCCTCCCC	TGTCACATAC	CAGGTGGGCT	ACGGACATCT	2040
CTGCACCTCG	TGAAACTGGA	CAAGAAGCCA	GTGGAGACCA	ACGGTCCTGA	TGCCAATCAC	2100
GAGTACAGTA	CAAGCAGCCA	GCATATGGAC	TACAAAGACA	CATTCGACAT	GCTGGACTCA	2160
CTGTTAAGTG	CCCAAGGAAT	GAACATGTAA				2190

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 729 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Met Ala Ser Gly Arg Gly Ala Ser Ser Arg Trp Phe Phe Thr Arg Glu
1 5 10 15

Gln Leu Glu Asn Thr Pro Ser Arg Arg Cys Gly Val Glu Ala Asp Lys
20 25 30

Glu Leu Ser Cys Arg Gln Gln Ala Ala Asn Leu Ile Gln Glu Met Gly
35 40 45

Gln Arg Leu Asn Val Ser Gln Leu Thr Ile Asn Thr Ala Ile Val Tyr
50 55 60

Met His Arg Phe Tyr Met His His Ser Phe Thr Lys Phe Asn Lys Asn
65 70 75 80

Ile Ile Ser Ser Thr Ala Leu Phe Leu Ala Ala Lys Val Glu Glu Gln
85 90 95

Ala Arg Lys Leu Glu His Val Ile Lys Val Ala His Ala Cys Leu His
100 105 110

Pro Leu Glu Pro Leu Leu Asp Thr Lys Cys Asp Ala Tyr Leu Gln Gln
115 120 125

Thr Gln Glu Leu Val Ile Leu Glu Thr Ile Met Leu Gln Thr Leu Gly
130 135 140

Phe Glu Ile Thr Ile Glu His Pro His Thr Asp Val Val Lys Cys Thr
145 150 155 160

Gln Leu Val Arg Ala Ser Lys Asp Leu Ala Gln Thr Ser Tyr Phe Met
165 170 175

Ala Thr Asn Ser Leu His Leu Thr Thr Phe Cys Leu Gln Tyr Lys Pro
180 185 190

Thr Val Ile Ala Cys Val Cys Ile His Leu Ala Cys Lys Trp Ser Asn
195 200 205

Trp Glu Ile Pro Val Ser Thr Asp Gly Lys His Trp Trp Glu Tyr Val
210 215 220

Asp Pro Thr Val Thr Leu Glu Leu Leu Asp Glu Leu Thr His Glu Phe
225 230 235 240

Leu Gln Ile Leu Glu Lys Thr Pro Asn Arg Leu Lys Lys Ile Arg Asn
245 250 255

Trp Arg Ala Asn Gln Ala Ala Arg Lys Pro Lys Val Asp Gly Gln Val
260 265 270

03951188-201592

Ser Glu Thr Pro Leu Leu Gly Ser Ser Leu Val Gln Asn Ser Ile Leu
275 280 285

Val Asp Ser Val Thr Gly Val Pro Thr Asn Pro Ser Phe Gln Lys Pro
290 295 300

Ser Thr Ser Ala Phe Pro Ala Pro Val Pro Leu Asn Ser Gly Asn Ile
305 310 315 320

Ser Val Gln Asp Ser His Thr Ser Asp Asn Leu Ser Met Leu Ala Thr
325 330 335

Gly Met Pro Ser Thr Ser Tyr Gly Leu Ser Ser His Gln Glu Trp Pro
340 345 350

Gln His Gln Asp Ser Ala Arg Thr Glu Gln Leu Tyr Ser Gln Lys Gln
355 360 365

Glu Thr Ser Leu Ser Gly Ser Gln Tyr Asn Ile Asn Phe Gln Gln Gly
370 375 380

Pro Ser Ile Ser Leu His Ser Gly Leu His His Arg Pro Asp Lys Ile
385 390 395 400

Ser Asp His Ser Ser Val Lys Gln Glu Tyr Thr His Lys Ala Gly Ser
405 410 415

Ser Lys His His Gly Pro Ile Ser Thr Thr Pro Gly Ile Ile Pro Gln
420 425 430

Lys Met Ser Leu Asp Lys Tyr Arg Glu Lys Arg Lys Leu Glu Thr Leu
435 440 445

Asp Leu Asp Val Arg Asp His Tyr Ile Ala Ala Gln Val Glu Gln Gln
450 455 460

His Lys Gln Gly Gln Ser Gln Ala Ala Ser Ser Ser Ser Val Thr Ser
465 470 475 480

Pro Ile Lys Met Lys Ile Pro Ile Ala Asn Thr Glu Lys Tyr Met Ala
485 490 495

Asp Lys Lys Glu Lys Ser Gly Ser Leu Lys Leu Arg Ile Pro Ile Pro
500 505 510

Pro Thr Asp Lys Ser Ala Ser Lys Glu Glu Leu Lys Met Lys Ile Lys
515 520 525

Val Ser Ser Ser Glu Arg His Ser Ser Ser Asp Glu Gly Ser Gly Lys
530 535 540

Ser Lys His Ser Ser Pro His Ile Ser Arg Asp His Lys Glu Lys His
545 550 555 560

Lys Glu His Pro Ser Ser Arg His His Thr Ser Ser His Lys His Ser
 565 570 575
 His Ser His Ser Gly Ser Ser Ser Gly Gly Ser Lys His Ser Ala Asp
 580 585 590
 Gly Ile Pro Pro Thr Val Leu Arg Ser Pro Val Gly Leu Ser Ser Asp
 595 600 605
 Gly Ile Ser Ser Ser Ser Ser Arg Lys Arg Leu His Val Asn
 610 615 620
 Asp Ala Ser His Asn His His Ser Lys Met Ser Lys Ser Ser Lys Ser
 625 630 635 640
 Ser Gly Ser Ser Ser Ser Ser Ser Val Lys Gln Tyr Ile Ser
 645 650 655
 Ser His Asn Ser Val Phe Asn His Pro Leu Pro Leu Leu Pro Cys His
 660 665 670
 Ile Pro Gly Gly Leu Arg Thr Ser Gln His Leu Val Lys Leu Asp Lys
 675 680 685
 Lys Pro Val Glu Thr Asn Gly Pro Asp Ala Asn His Glu Tyr Ser Thr
 690 695 700
 Ser Ser Gln His Met Asp Tyr Lys Asp Thr Phe Asp Met Leu Asp Ser
 705 710 715 720
 Leu Leu Ser Ala Gln Gly Met Asn Met
 725

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2360 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GGAAAGTGCCT GCAACCTTCG CCGCTGCCTT CTGGTTGAAG CACTATGGAG GGAGAGAGGA	60
AGAACACCAA CAAACGGTGG TATTCACTC GAGAACAGCT GGAAAATAGC CCATCCCGTC	120
GTTTTGGCGT GGACCCAGAT AAAGAACTTT CTTATGCCA GCAGGCAGGCC AATCTGCTTC	180
AGGACATGGG GCAGCGTCTT AACGTCTCAC AATTGACTAT CAACACTGCT ATAGTATACA	240
TGCATCGATT CTACATGATT CAGTCCTTCA CACGGTTCCC TGGAAATTCT GTGGCTCCAG	300

CAGCCTGTT	TCTAGCAGCT	AAAGTGGAGG	AGCAGCCAA	AAAATTGGAA	CATGTCATCA	360
AGGTAGCACA	TACTTGTCTC	CATCCTCAGG	AATCCCTTCC	TGATACTAGA	AGTGAGGCTT	420
ATTTGCAACA	AGTTCAAGAT	CTGGTCATT	TAGAAAGCAT	AATTTGCAG	ACTTTAGGCT	480
TTGAACTAAC	AATTGATCAC	CCACATACTC	ATGTAGTAAA	GTGCACTCAA	CTTGTTCGAG	540
CAAGCAAGGA	CTTAGCACAG	ACTTCTTACT	TCATGGCAAC	CAACAGCCTG	CATTTGACCA	600
CATTTAGCCT	GCAGTACACA	CCTCCTGTGG	TGGCCTGTGT	CTGCATTAC	CTGGCTTGCA	660
AGTGGTCCAA	TTGGGAGATC	CCAGTCTCAA	CTGACGGGAA	GCAC	ACTGGTGG	720
ACGCCACTGT	GACCTGGAA	CTTTTAGATG	AACTGACACA	TGAGTTCTA	CAGATTTGG	780
AGAAAAACTCC	CAACAGGCTC	AAACGCATT	GGAATTGGAG	GGCATGCGAG	GCTGCCAAGA	840
AAACAAAAGC	AGATGACCGA	GGAACAGATG	AAAAGACTTC	AGAGCAGACA	ATCCTCAATA	900
TGATTTCCA	GAGCTCTTCA	GACACAACCA	TTGCAGGTTT	AATGAGCATG	TCAACTTCTA	960
CCACAAAGTGC	AGTGCCTTCC	CTGCCAGTCT	CCGAAGAGTC	ATCCAGCAAC	TTAACCAAGTG	1020
TGGAGATGTT	GCCGGGCAAG	CGTTGGCTGT	CCTCCCAACC	TTCTTCAAA	CTAGAACCTA	1080
CTCAGGGTCA	TCGGACTAGT	GAGAATTAG	CACTTACAGG	AGTTGATCAT	TCCTTACAC	1140
AGGATGGTTC	AAATGCATT	ATTTCCCAGA	AGCAGAATAG	TAAGAGTGTG	CCATCAGCTA	1200
AAGTGTCACT	GAAAGAATAC	CGCGCGAACG	ATGCAGAAGA	ATTGGCTGCC	CAGAAGAGGC	1260
AACTGGAGAA	CATGGAAGCC	AATGTGAAGT	CACAATATGC	ATATGCTGCC	CAGAATCTCC	1320
TTTCTCATCA	TGATAGCCAT	TCTTCAGTCA	TTCTAAAAAT	GCCCATAGAG	GGTCAGAAA	1380
ACCCCGAGCG	GCCTTTCTG	GAAAAGGCTG	ACAAAACAGC	TCTAAAATG	AGAATCCCAG	1440
TGGCAGGTGG	AGATAAAGCT	GCGTCTCAA	AACCAGAGGA	GATAAAAATG	CGCATAAAAG	1500
TCCATGCTGC	AGCTGATAAG	CACAATTCTG	TAGAGGACAG	TGTTACAAAG	AGCCGAGAGC	1560
ACAAAGAAGA	GCGCAAGACT	CACCCATCTA	ATCATCATCA	TCATCATAAT	CACCACTCAC	1620
ACAAGCACTC	TCATTCCAA	CTTCCAGTTG	GTACTGGAA	CAAACGTCC	GGTGATCCAA	1680
AACATAGTAG	CCAGACAAGC	AACTTAGCAC	ATAAAACCTA	TAGCTTGTCT	AGTTCTTTT	1740
CCTCTTCCAG	TTCTACTCGT	AAAAGGGGAC	CCTCTGAAGA	GA	CTGGAGGG	1800
ATCATCCAGC	CAAGATTGCC	AAGAGTACTA	AATCCTCTTC	CCTAAATTTC	TCCTTCCCTT	1860
CACTTCCTAC	AATGGGTCAG	ATGCCTGGC	ATAGCTCAGA	CACAAGTGGC	CTTCCCTTT	1920

CACAGCCCAG	CTGTAAA	ACT CGTGC	CCCTC	ATT CGAA	ACT GGATA	AAAGGG CCC	ACT GGGG	1980
CCAATGGTCA	CAACACGACC	CAGACAATAG	ACTATCAAGA	CACTGTGAAT	ATGCTTC	ACT	2040	
CCCTGCTCAG	TGCC	CAGGGT GTTCAG	CCCA CTCAG	CCCAC TGCATTGAA	TTTGTTC	CGTC	2100	
CTTATAGTGA	CTATCTGAAT	CCTCGGTCTG	GTGGAATCTC	CTCGAGATCT	GGCAATACAG		2160	
ACAAACCCCG	GCCACCACCT	CTGCCATCAG	AACCTCCTCC	ACCACTTCCA	CCCCTTC	CTA	2220	
AGTAAAAAAA	GAAAAAGAAG	AGGAGAAAAA	AACTTCTTA	AAAAAACACA	TAATTTTCT		2280	
TTTTTTTTG	GGGAAAAAAA	AATTTTTTT	AAAATTTTT	CCCCAAGGGA	CGGGGGAAAA		2340	
TTTTATTTT	AAAATTTTT						2360	

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2181 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

ATGGAGGGAG	AGAGGAAGAA	CAACAACAAA	CGGTGGTATT	TCACTCGAGA	ACAGCTGGAA		60
AATAGCCCAT	CCC	GTGTTT	TGGCGTGGAC	CCAGATAAAG	AACTTTCTTA	TCGCCAGCAG	120
GCGGCCAATC	TGCTTCAGGA	CATGGGGCAG	CGTCTTAACG	TCTCACAAATT	GACTATCAAC		180
ACTGCTATAG	TATACATGCA	TCGATTCTAC	ATGATTCA	GTGATTCA	CCTTCACACG	GTTCCCTGGA	240
AATTCTGTGG	CTCCAGCAGC	CTTGT	TTCTA	GCAGCTAAAG	TGGAGGAGCA	GCCCA	300
TTGGAACATG	TCATCAAGGT	AGCACATACT	TGTCTCC	ATC	CTCAGGAATC	CCTTC	360
ACTAGAAGTG	AGGCTT	ATTT	GCAACAAGTT	CAAGATCTGG	TCATTTAGA	AAGCATAATT	420
TTGCAGACTT	TAGGCTT	GA	ACTAACAAATT	GATCACCCAC	ATACTCATGT	AGTAAAGTGC	480
ACTCAACTTG	TTCGAGCAAG	CAAGGACTTA	GCACAGACTT	CTTACTTCAT	GGCAACCAAC		540
AGCCTGCATT	TGACCACATT	TAGCCTGCAG	TACACACCTC	CTGTGGTGGC	CTGTGTCTGC		600
ATTCACCTGG	CTTGCAAGTG	GTCCAATTGG	GAGATCCCAG	TCTCAACTGA	CGGGAA	AGCAC	660
TGGTGGGAGT	ATGTTGACGC	CACTGTGACC	TTGGA	ACTTT	TAGATGA	ACT GACACATGAG	720
TTTCTACAGA	TTTGAGAA	AACTCC	AAAC	AGGCTCAAAC	GCATTG	GA TTGGAGGGCA	780

2025 RELEASE UNDER E.O. 14176

TGCGAGGCTG	CCAAGAAAAC	AAAAGCAGAT	GACCGAGGAA	CAGATGAAAA	GACTTCAGAG	840
CAGACAATCC	TCAATATGAT	TTCCCAGAGC	TCTTCAGACA	CAACCATTGC	AGGTTTAATG	900
AGCATGTCAA	CTTCTACCAAC	AAGTGCAGTG	CCTTCCCTGC	CAGTCTCCGA	AGAGTCATCC	960
AGCAACTTAA	CCAGTGTGGA	GATGTTGCCG	GGCAAGCGTT	GGCTGTCCTC	CCAACCTTCT	1020
TTCAAACTAG	AACCTACTCA	GGGTCATCGG	ACTAGTGAGA	ATTTAGCACT	TACAGGAGTT	1080
GATCATTCCCT	TACCACAGGA	TGGTTCAAAT	GCATTTATTT	CCCAGAAGCA	GAATAGTAAG	1140
AGTGTGCCAT	CAGCTAAAGT	GTCACTGAAA	GAATACCGCG	CGAAGCATGC	AGAAGAATTG	1200
GCTGCCAGA	AGAGGCAACT	GGAGAACATG	GAAGCCAATG	TGAAGTCACA	ATATGCATAT	1260
GCTGCCAGA	ATCTCCTTTC	TCATCATGAT	AGCCATTCTT	CAGTCATTCT	AAAAATGCC	1320
ATAGAGGGTT	CAGAAAACCC	CGAGCGGCCT	TTTCTGGAAA	AGGCTGACAA	AACAGCTCTC	1380
AAAATGAGAA	TCCCAGTGGC	AGGTGGAGAT	AAAGCTGCGT	CTTCAAAACC	AGAGGAGATA	1440
AAAATGCGCA	TAAAAGTCCA	TGCTGCAGCT	GATAAGCACA	ATTCTGTAGA	GGACAGTGTT	1500
ACAAAGAGCC	GAGAGCACAA	AGAAGAGCGC	AAGACTCACC	CATCTAATCA	TCATCATCAT	1560
CATAATCACC	ACTCACACAA	GCACCTCAT	TCCCAACTTC	CAGTTGGTAC	TGGGAACAAA	1620
CGTCCTGGTG	ATCCAAAACA	TAGTAGCCAG	ACAAGCAACT	TAGCACATAA	AACCTATAGC	1680
TTGTCTAGTT	CTTTTCCTC	TTCCAGTTCT	ACTCGTAAA	GGGGACCCCTC	TGAAGAGACT	1740
GGAGGGGCTG	TGTTTGATCA	TCCAGCCAAG	ATTGCCAAGA	GTACTAAATC	CTCTTCCCTA	1800
AATTTCTCCT	TCCCTTCACT	TCCTACAATG	GGTCAGATGC	CTGGGCATAG	CTCAGACACA	1860
AGTGGCCTTT	CCTTTTCACA	GCCCAGCTGT	AAAACCTCGTG	TCCCTCATTC	GAAACTGGAT	1920
AAAGGGCCCA	CTGGGGCCAA	TGGTCACAAC	ACGACCCAGA	CAATAGACTA	TCAAGACACT	1980
GTGAATATGC	TTCACTCCCT	GCTCAGTGCC	CAGGGTGTTC	AGCCCACCTCA	GCCCAC TGCA	2040
TTTGAATTTG	TTCGTCCTTA	TAGTGACTAT	CTGAATCCTC	GGTCTGGTGG	AATCTCCTCG	2100
AGATCTGGCA	ATACAGACAA	ACCCCGGCCA	CCACCTCTGC	CATCAGAACCC	TCCTCCACCA	2160
CTTCCACCCC	TTCTTAAGTA	A				2181

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 726 amino acids
- (B) TYPE: amino acid

(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Met Glu Gly Glu Arg Lys Asn Asn Asn Lys Arg Trp Tyr Phe Thr Arg
1 5 10 15

Glu Gln Leu Glu Asn Ser Pro Ser Arg Arg Phe Gly Val Asp Pro Asp
20 25 30

Lys Glu Leu Ser Tyr Arg Gln Gln Ala Ala Asn Leu Leu Gln Asp Met
35 40 45

Gly Gln Arg Leu Asn Val Ser Gln Leu Thr Ile Asn Thr Ala Ile Val
50 55 60

Tyr Met His Arg Phe Tyr Met Ile Gln Ser Phe Thr Arg Phe Pro Gly
65 70 75 80

Asn Ser Val Ala Pro Ala Ala Leu Phe Leu Ala Ala Lys Val Glu Glu
85 90 95

Gln Pro Lys Lys Leu Glu His Val Ile Lys Val Ala His Thr Cys Leu
100 105 110

His Pro Gln Glu Ser Leu Pro Asp Thr Arg Ser Glu Ala Tyr Leu Gln
115 120 125

Gln Val Gln Asp Leu Val Ile Leu Glu Ser Ile Ile Leu Gln Thr Leu
130 135 140

Gly Phe Glu Leu Thr Ile Asp His Pro His Thr His Val Val Lys Cys
145 150 155 160

Thr Gln Leu Val Arg Ala Ser Lys Asp Leu Ala Gln Thr Ser Tyr Phe
165 170 175

Met Ala Thr Asn Ser Leu His Leu Thr Thr Phe Ser Leu Gln Tyr Thr
180 185 190

Pro Pro Val Val Ala Cys Val Cys Ile His Leu Ala Cys Lys Trp Ser
195 200 205

Asn Trp Glu Ile Pro Val Ser Thr Asp Gly Lys His Trp Trp Glu Tyr
210 215 220

Val Asp Ala Thr Val Thr Leu Glu Leu Leu Asp Glu Leu Thr His Glu
225 230 235 240

Phe Leu Gln Ile Leu Glu Lys Thr Pro Asn Arg Leu Lys Arg Ile Trp
245 250 255

Asn Trp Arg Ala Cys Glu Ala Ala Lys Lys Thr Lys Ala Asp Asp Arg
260 265 270

Gly Thr Asp Glu Lys Thr Ser Glu Gln Thr Ile Leu Asn Met Ile Ser
275 280 285

Gln Ser Ser Ser Asp Thr Thr Ile Ala Gly Leu Met Ser Met Ser Thr
290 295 300

Ser Thr Thr Ser Ala Val Pro Ser Leu Pro Val Ser Glu Glu Ser Ser
305 310 315 320

Ser Asn Leu Thr Ser Val Glu Met Leu Pro Gly Lys Arg Trp Leu Ser
325 330 335

Ser Gln Pro Ser Phe Lys Leu Glu Pro Thr Gln Gly His Arg Thr Ser
340 345 350

Glu Asn Leu Ala Leu Thr Gly Val Asp His Ser Leu Pro Gln Asp Gly
355 360 365

Ser Asn Ala Phe Ile Ser Gln Lys Gln Asn Ser Lys Ser Val Pro Ser
370 375 380

Ala Lys Val Ser Leu Lys Glu Tyr Arg Ala Lys His Ala Glu Glu Leu
385 390 395 400

Ala Ala Gln Lys Arg Gln Leu Glu Asn Met Glu Ala Asn Val Lys Ser
405 410 415

Gln Tyr Ala Tyr Ala Ala Gln Asn Leu Leu Ser His His Asp Ser His
420 425 430

Ser Ser Val Ile Leu Lys Met Pro Ile Glu Gly Ser Glu Asn Pro Glu
435 440 445

Arg Pro Phe Leu Glu Lys Ala Asp Lys Thr Ala Leu Lys Met Arg Ile
450 455 460

Pro Val Ala Gly Gly Asp Lys Ala Ala Ser Ser Lys Pro Glu Glu Ile
465 470 475 480

Lys Met Arg Ile Lys Val His Ala Ala Asp Lys His Asn Ser Val
485 490 495

Glu Asp Ser Val Thr Lys Ser Arg Glu His Lys Glu Glu Arg Lys Thr
500 505 510

His Pro Ser Asn His His His His Asn His His Ser His Lys His
515 520 525

Ser His Ser Gln Leu Pro Val Gly Thr Gly Asn Lys Arg Pro Gly Asp
530 535 540

Pro Lys His Ser Ser Gln Thr Ser Asn Leu Ala His Lys Thr Tyr Ser
 545 550 555 560
 Leu Ser Ser Ser Phe Ser Ser Ser Ser Thr Arg Lys Arg Gly Pro
 565 570 575
 Ser Glu Glu Thr Gly Gly Ala Val Phe Asp His Pro Ala Lys Ile Ala
 580 585 590
 Lys Ser Thr Lys Ser Ser Ser Leu Asn Phe Ser Phe Pro Ser Leu Pro
 595 600 605
 Thr Met Gly Gln Met Pro Gly His Ser Ser Asp Thr Ser Gly Leu Ser
 610 615 620
 Phe Ser Gln Pro Ser Cys Lys Thr Arg Val Pro His Ser Lys Leu Asp
 625 630 635 640
 Lys Gly Pro Thr Gly Ala Asn Gly His Asn Thr Thr Gln Thr Ile Asp
 645 650 655
 Tyr Gln Asp Thr Val Asn Met Leu His Ser Leu Leu Ser Ala Gln Gly
 660 665 670
 Val Gln Pro Thr Gln Pro Thr Ala Phe Glu Phe Val Arg Pro Tyr Ser
 675 680 685
 Asp Tyr Leu Asn Pro Arg Ser Gly Gly Ile Ser Ser Arg Ser Gly Asn
 690 695 700
 Thr Asp Lys Pro Arg Pro Pro Pro Leu Pro Ser Glu Pro Pro Pro Pro
 705 710 715 720
 Leu Pro Pro Leu Pro Lys
 725

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TTCCCCACCAA TGCTTTCC

18

(2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CCATCAGTTG ATACAGGGAT CT

22

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGAATTCAAGA AGGTTGTAAG ATGC

24

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

ACACACAGAT GTGGTGAAAT GTACCCA

27

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GCATCTTACA ACCTTCTG

18

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGAATTCAATG GAAAGCATTG GTGGGAAT

28

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CCTCCACTAC TGGTTTGCCT GG

22

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GGACTAGTAT AAATATGGCG TCGGGCCGTG

30

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GGAGATCTTA CATGTTCAATT CCTTGGG

27

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GGAGACAAAGT ATGTGCTACC TTGATGACA

29

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GGAATTCGGG CTGCTCCTCC ACTTTAG

27

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GGAATTCGCT GCTGGAGCCA CAGAA

25

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GTGTCACTGA AAGAATACCG

20

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GGAATTCAAGG TGGAGATAAA GCTGC

25

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GCTCTAGATA AATATGGAGG GAGAGAGGAA

30

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GGAATTCTTA CTTAGGAAGG GGTGGAAGTG

30

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GGAATTCTTA CTTAGGAAGG GGTGGAAGTG GTGGAGGAGG TTAC

44

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Ala Cys Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro
1 5 10 15

Ser Tyr Ser Pro Thr Ser Pro Ser Lys Lys
20 25

08954188 - 201597